



資料分析專題報告

主題：台北市100年至109年
人口變化分析



2021/02/01

目錄

CONTENTS

01



選題動機

02



操作流程

03



結果說明

04



作業心得

PART ONE

01

Annual work summary.

選題動機



一則新聞引發注意

財經 熱線

台灣人口負成長「脫北」加劇 台北市去年淨遷出4.44萬人

記者：工商時報郭及天 | 2021-01-11 21:17

負成長 脫北 台南 桃園 房價 台北市



(圖／報系資料庫)

A+ A-

台灣人口負成長趨勢確立，2020首度出現負成長，房價最高、「脫北風潮」加劇的台北市，人口流失最嚴重，去年底台北市人口260.24萬人，總人口數創下23年來最低，一年間減少4.26萬人，人口減少數量更創下27年來最高。

過去十年北市人口如何變化？

- 整體變化？
- 區域變化？

台北人口年齡分布如何？

- 年齡層分布？老化嚴重？
- 老年及新生兒變化？

現況

- 男女占比差異？
- 明年？

透過年份、年齡、區域、性別四個維度來了解台北市人口的變化

資料來源

- <http://163.29.157.32:8080/dataset/16e8d284-6121-4646-b6d5-72166175fc0c>
- 臺北市各區人口年齡分配(單齡)按月別時間數列統計資料
- 臺北市政府主計處
- 包含100-109十年間台北市各區域0-100歲人口數 (區分男女)
- [20210120134809\(原始\).csv](#) (2565KB,4386行*125列)
- [Taipei_people_UTF8.csv](#) (實際使用428KB)
- 參考資料:[108年第37週內政統計通報_生命表_1.pdf](#)

二、平均壽命

- (一)全體：107年國人平均壽命為80.69歲，較106年增加0.30歲，創歷年新高。從長期趨勢觀察，隨著醫療水準提升、食品安全重視、生活品質提高及運動風氣盛行，國人平均壽命長期呈現上升趨勢，從97年78.57歲增至107年80.69歲，顯示國人已越來越長壽。
- (二)性別：107年我國男性平均壽命為77.55歲，女性為84.05歲，與106年比較，男性增加0.26歲、女性增加0.35歲。
- (三)直轄市：以臺北市83.63歲最高，其餘依序為新北市、桃園市、臺中市、臺南市及高雄市，男、女性亦以臺北市最高。六直轄市在全體、男性及女性平均壽命皆呈現自北而南遞減的情形。

PART ONE
02

Annual work summary.

操作流程



操作流程

索引	年份	月份	年月底別	行政區	性別	總計	0	0~9	10~19	20~29	30~39	40~49
0	100年	1月	100年1月	總計	總計	2622883	19043	227138	316044	337656	411385	434
1	100年	1月	100年1月	總計	男	1264359	9883	118260	164667	169528	189343	204
2	100年	1月	100年1月	總計	女	1358524	9160	108878	151377	168128	222042	233
3	100年	1月	100年1月	松山區	總計	208769	1451	18976	26655	23559	31007	353
4	100年	1月	100年1月	松山區	男	98992	773	9862	13972	11716	13638	150
5	100年	1月	100年1月	松山區	女	109777	678	9114	12683	11843	17369	199
6	100年	1月	100年1月	信義區	總計	225375	1678	18250	24394	29861	36057	364
7	100年	1月	100年1月	信義區	男	108498	863	9529	12737	14905	16604	169
8	100年	1月	100年1月	信義區	女	116877	815	8721	11657	14956	19453	195
9	100年	1月	100年1月	大安區	總計	311896	2212	28222	38299	33836	44215	513

整理資料

```
import pandas as pd
import csv
import matplotlib.pyplot as plt
import matplotlib.font_manager as fm
plt.rcParams['font.family']='Microsoft YaHei'
#=====
df=pd.read_csv('Taipei_people_UTF8.csv',encoding="UTF8")
#=====check if exist NAN =====
print(df.isnull().values.any()) #檢查整個DF 有沒有NAN
#print(df.isna().sum().sum()) #計算整個DF 中共有幾個NAN
#print(df.isnull())
df.loc[:, '總計':].astype(int) #change some columns into INT data type
df.insert(0, column='年份', value="ok") #在DF中新增一欄'年份' 預設值=ok
for i in range(len(df)): #將年份設為'年月底別' 的前三位
    df.loc[i, '年份']=df.loc[i, "年月底別"][0:4]
df.insert(1, column='月份', value="ok") #在DF中新增一欄'月份' 預設值=ok
for i in range(len(df)): #將月份設為'年月底別' 的第四位後
    df.loc[i, '月份']=df.loc[i, "年月底別"][4:]
#=====
```

設定條件

- 整理原始資料 (瘦身)
- Pandas 讀取CSV 轉為 DataFrame (astype(int))
- 觀察了解行列結構

- 設定MASK
- 利用df.loc(切片)、df.index、df.column、df.values抓取所需資料

視覺輸出

- Matplotlib
- bar/pie
- savefig

設定條件

```

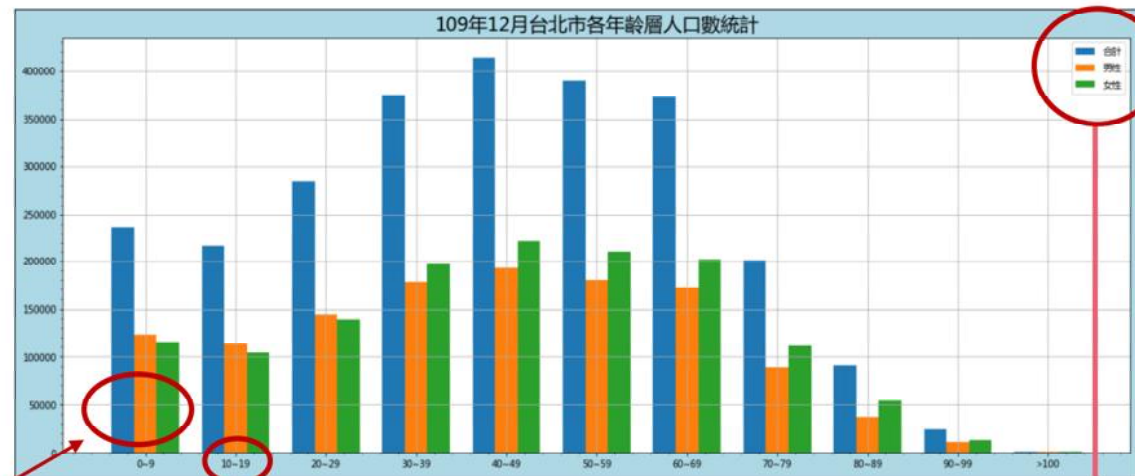
144 mask6=(df['性別']=="總計")&((df['年份']=="100年")^(df['年份']=="109年"))&(df['月份']=="12月")&(df['行政區']!="總計")
145 #把109年12月及100年12月的分區人口數相減,看其變化
146 df6=df[mask6]
147 outlist6=[]
148 arealist=list(df6['行政區'].head(12).values)
149 for item in (df6.head(12)).index: #抓前面12筆資料的index
150     outlist6.append(df6.loc[(item+3990),'總計']-df6.loc[(item),'總計'])
151 #109年資料與100年資料行索引值相差3990/108-109相差468
    
```

Index相差
3990

索引	年份	月份	年月底別	行政區	性別	總計	0	0~9	10~19	20~29	30~39	40~49	50~59	60~69	70~79	80~89	90~99	>100
432	100年	12月	100年12月	松山區	總計	209948	1771	19227	26181	23031	31638	35179	32498	21962	11912	7003	1266	51
435	100年	12月	100年12月	信義區	總計	226541	2179	18581	23773	29035	37170	36055	36227	23738	12909	7722	1291	40
438	100年	12月	100年12月	大安區	總計	313846	2685	28829	37780	32980	45166	51521	49312	34658	19490	11682	2352	76
441	100年	12月	100年12月	中山區	總計	224102	2208	18311	22759	27367	37280	36845	37739	23008	13025	6725	1009	34
444	100年	12月	100年12月	中正區	總計	161449	1498	15737	21213	17623	24419	26524	23814	15607	9636	5631	1172	73
447	100年	12月	100年12月	大同區	總計	126640	1257	10672	13845	16194	21435	19884	19739	12439	8350	3444	600	38
450	100年	12月	100年12月	萬華區	總計	190855	1959	14735	17923	25261	32520	29249	30318	19490	13132	7111	1078	38
453	100年	12月	100年12月	文山區	總計	266442	2556	24501	33801	32756	42885	46501	40462	23017	12804	8361	1313	41
456	100年	12月	100年12月	南港區	總計	116131	1271	10501	12950	16093	21298	18471	17287	10372	5835	2901	417	6
459	100年	12月	100年12月	內湖區	總計	275652	2660	25403	37231	38678	44586	47536	45502	19684	10233	5981	804	14
462	100年	12月	100年12月	士林區	總計	287072	2548	23787	32473	38508	46316	44734	47272	28365	16043	8230	1300	44
465	100年	12月	100年12月	北投區	總計	252290	2460	22276	29961	33605	42574	40007	39549	23352	12873	6961	1094	38
4422	109年	12月	109年12月	松山區	總計	200316	1248	18934	18580	20383	26157	32224	28879	28189	17232	7454	2169	115
4425	109年	12月	109年12月	信義區	總計	215240	1544	17809	15335	23568	31732	33389	32562	32148	18594	7876	2133	94
4428	109年	12月	109年12月	大安區	總計	302644	1927	30122	28510	29530	37167	47461	43467	42865	27269	12506	3587	160
4431	109年	12月	109年12月	中山區	總計	223876	1562	19235	16001	23246	32668	35471	34930	34213	18060	8067	1922	63
4434	109年	12月	109年12月	中正區	總計	155397	1009	16521	15327	15043	20948	25813	21401	20612	12111	5978	1567	76
4437	109年	12月	109年12月	大同區	總計	123849	873	11044	9733	13186	18644	19957	18198	17772	9544	4831	885	55
4440	109年	12月	109年12月	萬華區	總計	183088	1230	13794	11771	20164	27394	28544	28575	27806	15334	7724	1912	70
4443	109年	12月	109年12月	文山區	總計	268130	1760	23927	24015	31665	37623	43075	42671	36156	18378	8095	2432	93
4446	109年	12月	109年12月	南港區	總計	118758	877	11036	9094	13738	18844	19903	17757	15807	8237	3532	778	32
4449	109年	12月	109年12月	內湖區	總計	282525	2152	27968	24619	34128	44723	44963	43298	39521	15195	6327	1735	48
4452	109年	12月	109年12月	士林區	總計	278451	1849	23697	21906	31321	41193	42465	41088	42196	21953	10078	2454	100
4455	109年	12月	109年12月	北投區	總計	250144	1798	22705	21082	27790	37084	40656	36660	35770	18384	7974	1955	84

數據輸出

```
32 #=====各年齡層分布=====
33
34 mask1=(df['行政區']=="總計")&(df['年月底別']=="109年12月")
35 df1=(df[mask1].loc[:, '0~9':])
36 #print(df1.index)
37 reslist=[] #用於存放109年12月三筆的各年齡層人數
38 for item in df1.index: #找到109/12三筆資料的索引值
39     reslist.append(list(df1.loc[item,:].values))
40 columnlist=list(df1.loc[:, '0~9':].columns)
41
42 plt.figure(figsize=(20,8),facecolor='lightblue')
43 x1=[p for p in range(1,12)]
44 width=0.25
45 Data_1=plt.bar(x1,reslist[0],width=0.25,label='合計')
46 x2=[p + width for p in x1]
47 Data_2=plt.bar(x2,reslist[1],width=0.25,label='男性')
48 x3=[p + width for p in x2]
49 Data_3=plt.bar(x3,reslist[2],width=0.25,label='女性')
50 plt.xticks(x2,columnlist) #將底標置中
51 plt.title('109年12月台北市各年齡層人口數統計',fontsize=20)
52 plt.legend(handles=[Data_1,Data_2,Data_3])
53 plt.grid()
54 plt.minorticks_on()
55 plt.savefig('total_109.png')
56 plt.show()
```



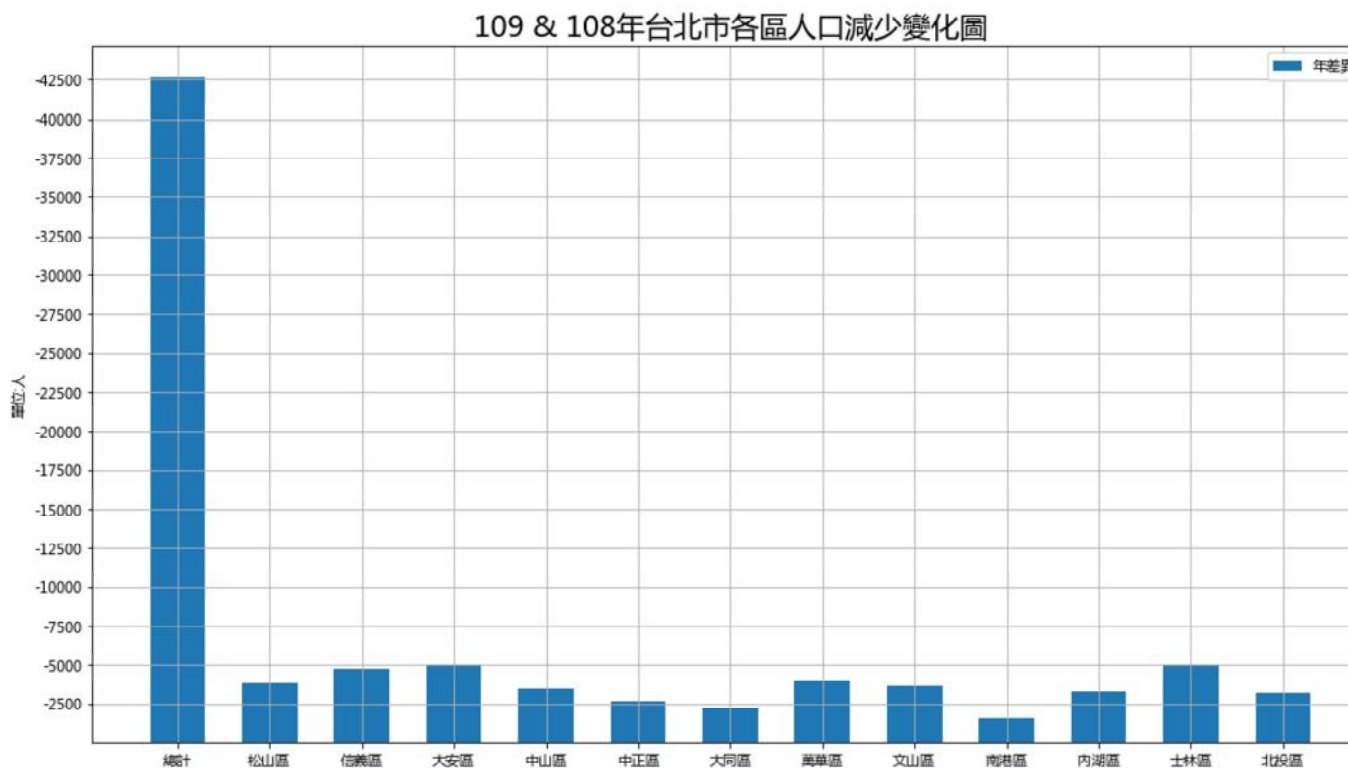
PART ONE
03

Annual work summary.

結果説明

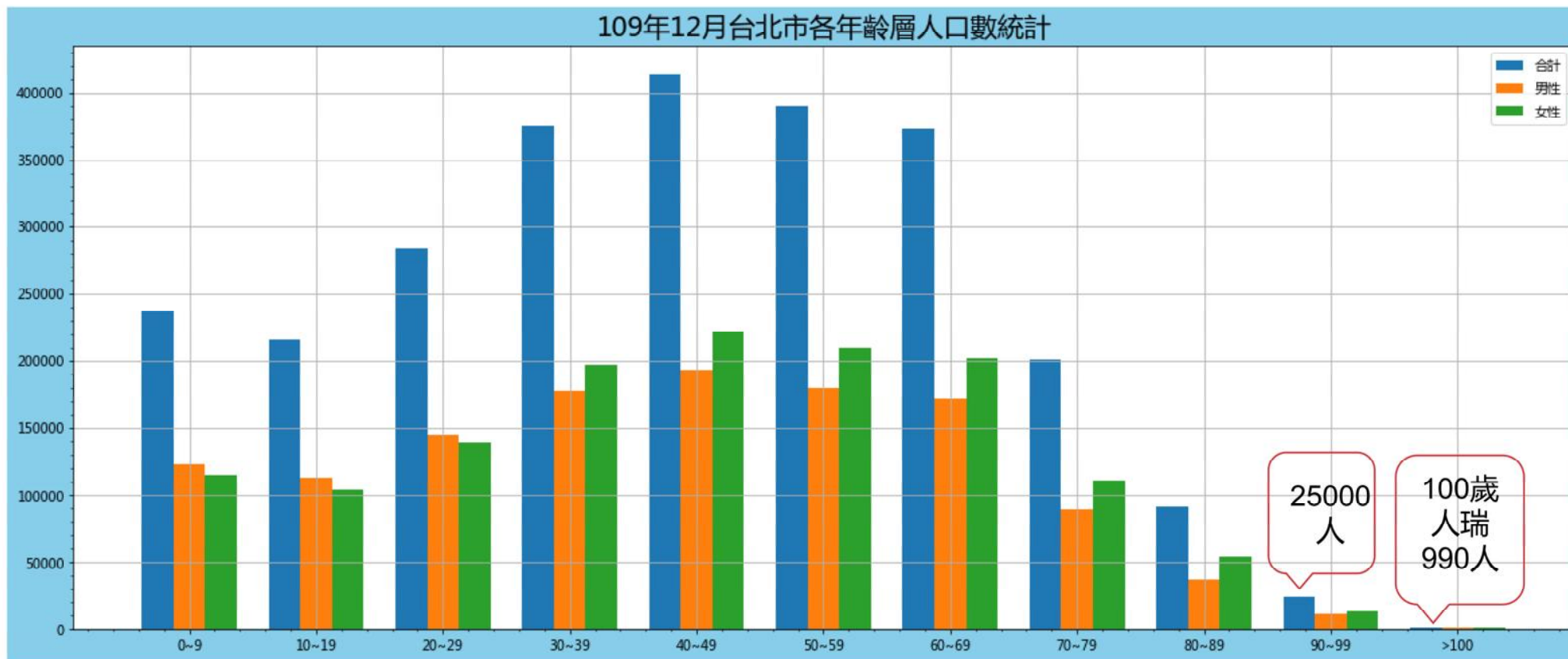


台北市人口減少警訊



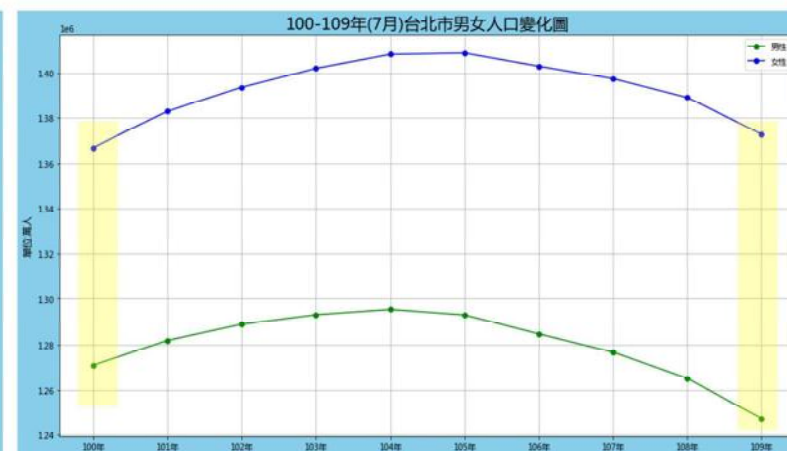
- 台北市2019年12月總人口數2,602,418人，相較108年12月2,645,041人減少42,623人
- PS:新北市2019/12月 401萬人

109年12月台北市人口概況



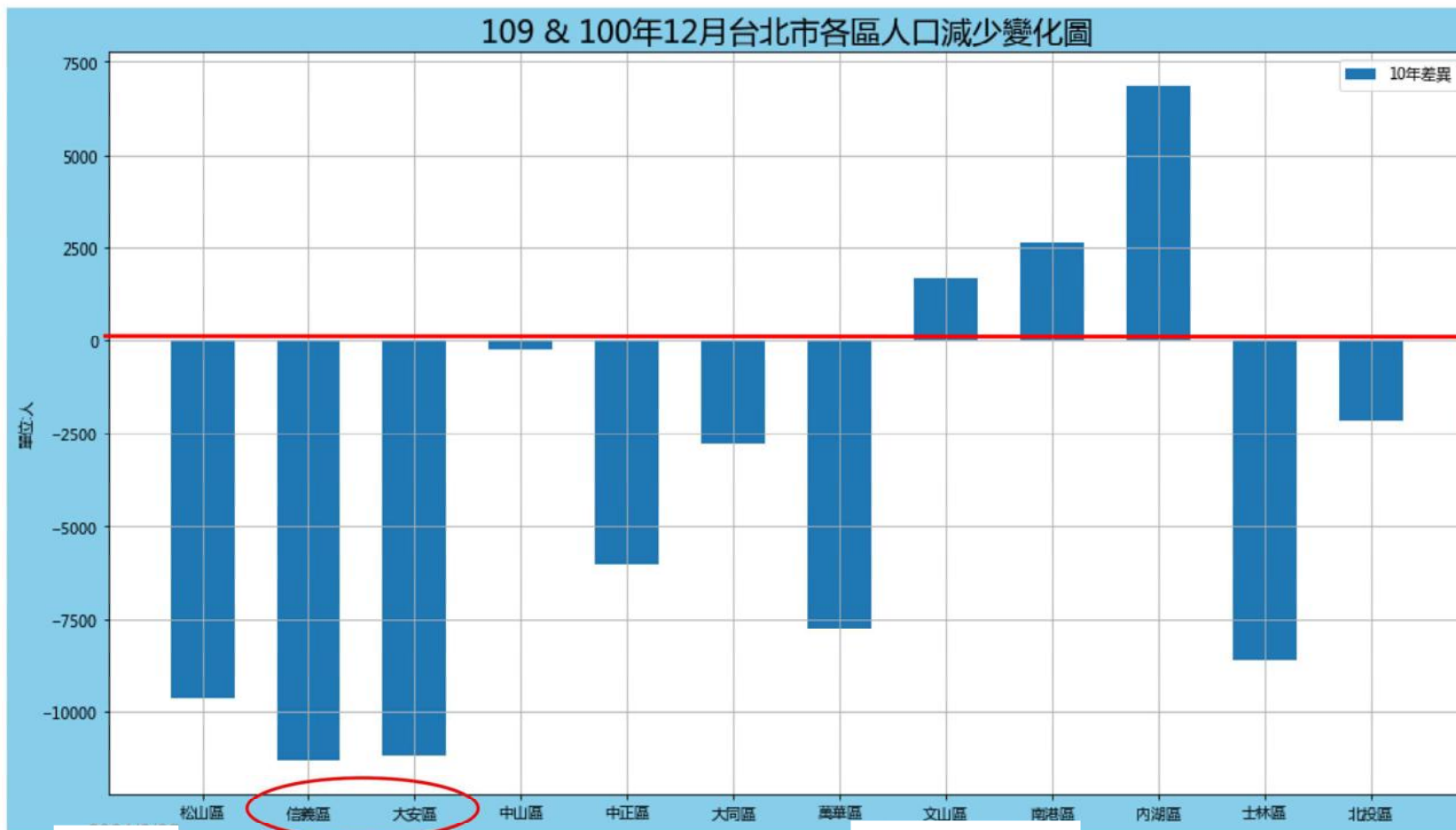
1. 人口數主要分佈在30~60歲之間
2. 30歲前男性人口數多於女性
3. 70歲以上人口佔總比11%

10年間台北市總人口數變化



1. 100年7月總人口2,637,866;109年7月為2,620,037, 減少17829人
2. 104年開始, 台北市人口總數呈現連續下滑趨勢
3. **10年中, 北市女性人口數始終大於男性, 且差異自100年98282人放大到109年的125184人**
4. 近三年女性佔比52.3%,男性47.7%

10(100-109)年間各行政區人口變化



1. 除文山/南港/內湖三行政區人口數有增加外，其餘均減少
2. 以信義及大安區減少最多

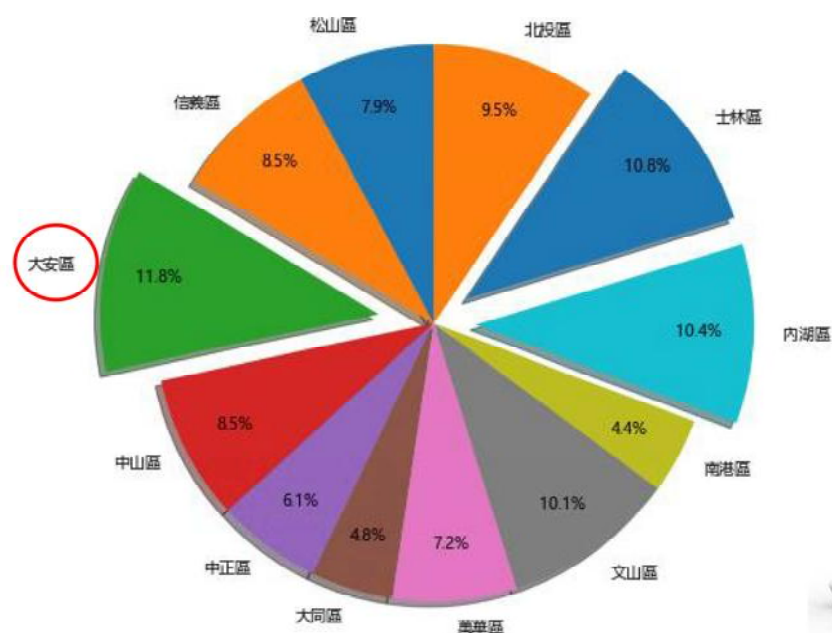
台北市各行政區房價排名		
排名	平均單價	行政區
1	93萬/坪	大安區
2	88.7萬/坪	信義區
3	82.8萬/坪	中正區
4	75.3萬/坪	松山區
5	71.5萬/坪	中山區
6	66萬/坪	士林區
7	65.1萬/坪	內湖區
8	65.1萬/坪	大同區
9	63.7萬/坪	南港區
10	54.9萬/坪	北投區
11	53.6萬/坪	萬華區
12	52.7萬/坪	文山區

統計時間：2019/6/1~2020/5/31

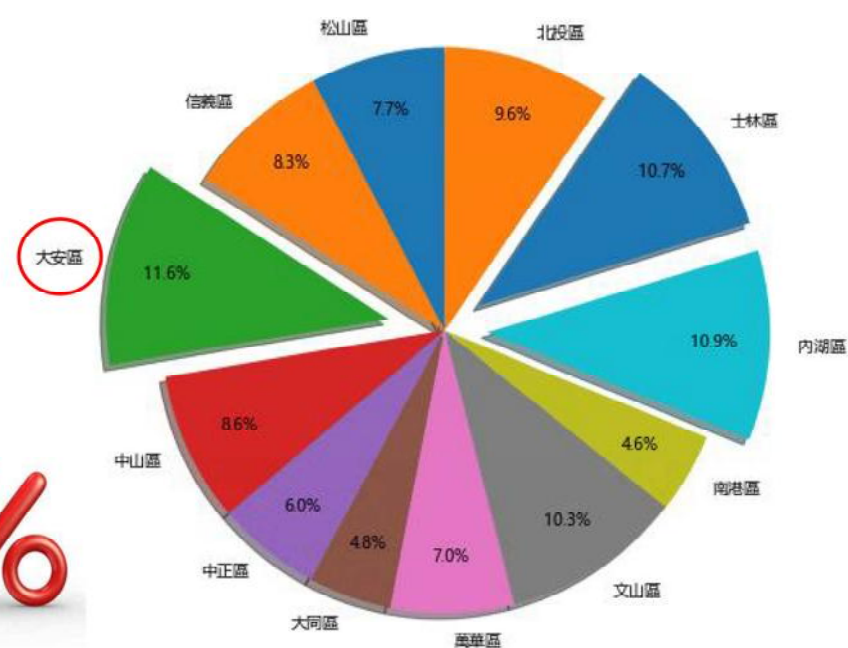
統計類別僅包含社區、大樓，排除一樓、地下室、特殊交易

100&109年各行政區人數佔比

100年台北市各區人口數佔比



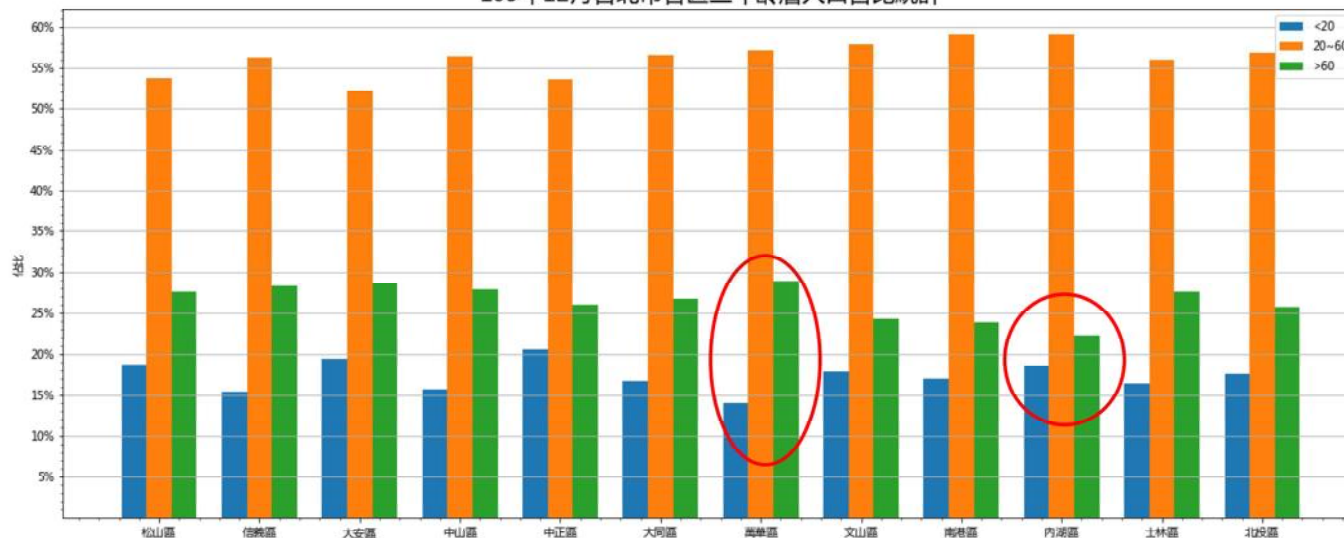
109年台北市各區人口數佔比



1. 10年間各行政區人口佔比沒有明顯變化
2. 人口佔比10%以上：大安區（最高）、文山/士林及內湖區
3. 人口佔比增加：內湖/文山/南港/中山/北投(人口往外圍移動)
4. 人口佔比減少：松山/信義/大安/中正/萬華/士林

109年12月各區年齡層佔比分析

109年12月台北市各區三年齡層人口占比統計



```

161 #=====109/12台北市各區三年齡層人口占比統計=====
162 mask7=(df['性別']=="總計")&(df['年月底別']=="109年12月")&(df['行政區']!="總計")
163 df7=df[mask7]
164 #agelist=[p-0.5 for p in range(5,110,10)]
165 ratio1=[]
166 ratio2=[]
167 ratio3=[]
168 result=[]
169 for item in df7.index:
170     tot=df7['總計'][item]
171     y019=df7.loc[item,'0~9':'10~19'].sum()
172     y2060=df7.loc[item,'20~29':'50~59'].sum()
173     above60=df7.loc[item,'60~69':].sum()
174     ratio1.append((y019/tot)*100)
175     ratio2.append((y2060/tot)*100)
176     ratio3.append((above60/tot)*100)
    
```

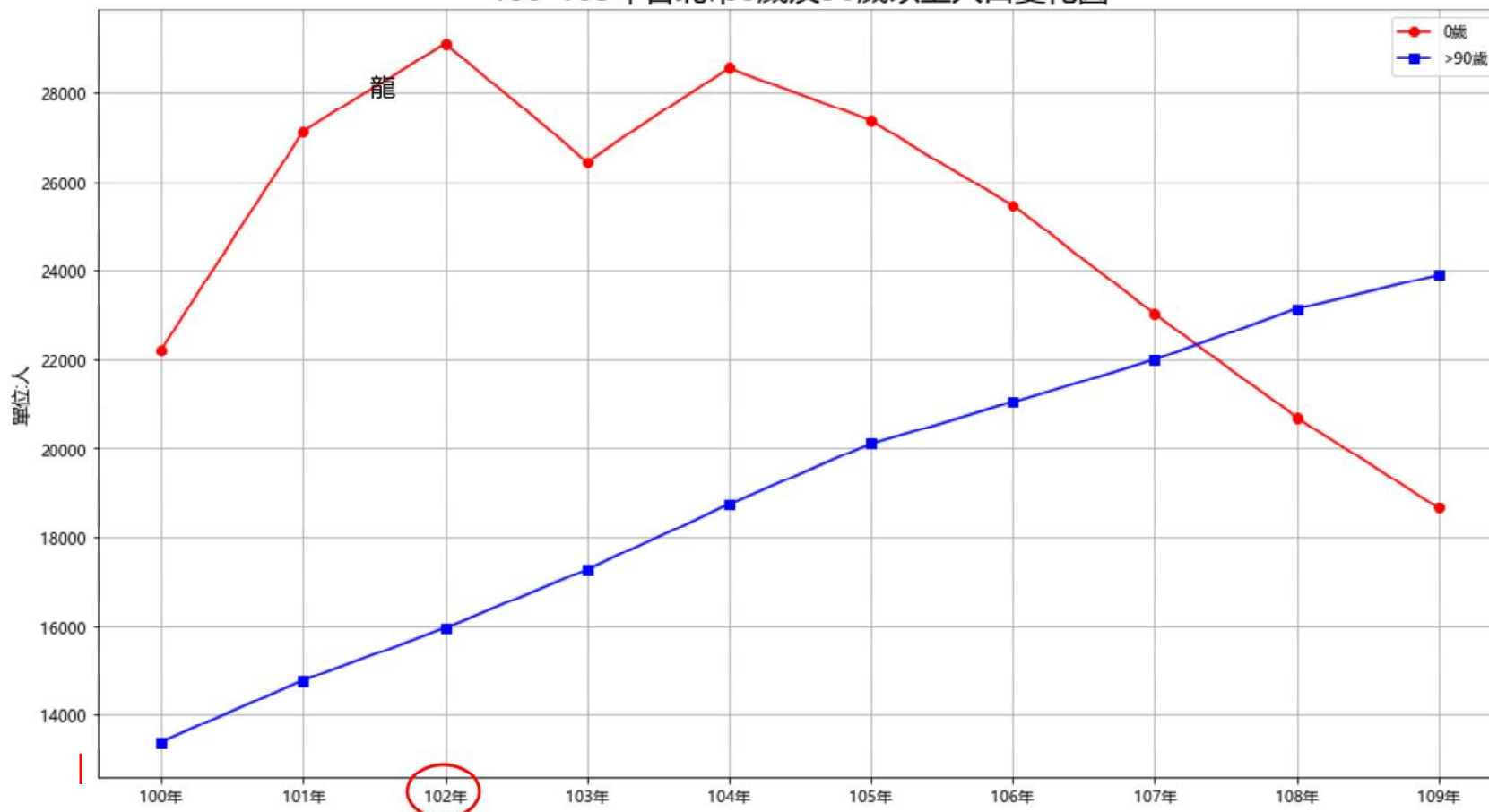
松山區	平均年齡: 43.5
信義區	平均年齡: 44.5
大安區	平均年齡: 43.8
中山區	平均年齡: 44.3
中正區	平均年齡: 42.4
大同區	平均年齡: 43.5
萬華區	平均年齡: 45.2
文山區	平均年齡: 42.5
南港區	平均年齡: 42.4
內湖區	平均年齡: 41.1
士林區	平均年齡: 43.8
北投區	平均年齡: 42.9

1. 將年齡分為<20、20-60、>60三個區塊來看各區中佔比差異
2. “20歲以下”佔比最高為中正區 (>20%)
3. “20-60歲”在南港及內湖佔比最高，接近60%
4. 萬華區老年人佔比最高，同時小於20歲的比例偏低
5. 內湖區相對“較年輕化”

台北市10年間0歲及90歲以上人口變化



100-109年台北市0歲及90歲以上人口變化圖



7月

PART ONE
04

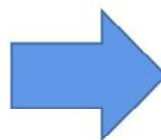
Annual work summary.

解決問題



遇到問題 (索引值)

原始 df - DataFrame								
索引	年份	月份	年月底別	行政區	性別	總計	0	0~9
0	100年	1月	100年1月	總計	總計	2622883	19043	227138
1	100年	1月	100年1月	總計	男	1264359	9883	118260
2	100年	1月	100年1月	總計	女	1358524	9160	108878
3	100年	1月	100年1月	松山區	總計	208769	1451	18976
4	100年	1月	100年1月	松山區	男	98992	773	9862
5	100年	1月	100年1月	松山區	女	109777	678	9114
6	100年	1月	100年1月	信義區	總計	225375	1678	18250
7	100年	1月	100年1月	信義區	男	108498	863	9529
8	100年	1月	100年1月	信義區	女	116877	815	8721
9	100年	1月	100年1月	大安區	總計	311896	2212	28222
10	100年	1月	100年1月	大安區	男	146858	1166	14676
11	100年	1月	100年1月	大安區	女	165038	1046	13546
12	100年	1月	100年1月	中山區	總計	220573	1591	17569
13	100年	1月	100年1月	中山區	男	103763	775	9070
14	100年	1月	100年1月	中山區	女	116810	816	8499
15	100年	1月	100年1月	中正區	總計	150769	1063	15219



篩選後 df6 - DataFrame								
索引	年份	月份	年月底別	行政區	性別	總計	0	0~9
432	100年	12月	100年12月	松山區	總計	209948	1771	19227
435	100年	12月	100年12月	信義區	總計	226541	2179	18581
438	100年	12月	100年12月	大安區	總計	313846	2685	28829
441	100年	12月	100年12月	中山區	總計	224102	2208	18311
444	100年	12月	100年12月	中正區	總計	161449	1498	15737
447	100年	12月	100年12月	大同區	總計	126640	1257	10672
450	100年	12月	100年12月	萬華區	總計	190855	1959	14735
453	100年	12月	100年12月	文山區	總計	266442	2556	24501
456	100年	12月	100年12月	南港區	總計	116131	1271	10501
459	100年	12月	100年12月	內湖區	總計	275652	2660	25403
462	100年	12月	100年12月	士林區	總計	287072	2548	23787
465	100年	12月	100年12月	北投區	總計	252290	2460	22276
4422	109年	12月	109年12月	松山區	總計	200316	1248	18934
4425	109年	12月	109年12月	信義區	總計	215240	1544	17809
4428	109年	12月	109年12月	大安區	總計	302644	1927	30122
4431	109年	12月	109年12月	中山區	總計	222876	1563	18335

1. 經過Mask 處理後的DataFrame 並不是一個新的DF,仍用原來的行Index
2. 處理迴圈的過程中大多使用 "for item in df["某欄位"]"
3. 使用 df.loc 未使用 df.iloc

遇到問題 (Series轉Dataframe)

temp Series (10,) Series object of pandas.core.series module

索引	總計
231	nan
699	2431
1149	1828
1617	1432
2085	1401
2349	-14
2817	-1003
3285	-1154
3753	-1474
4221	-2414

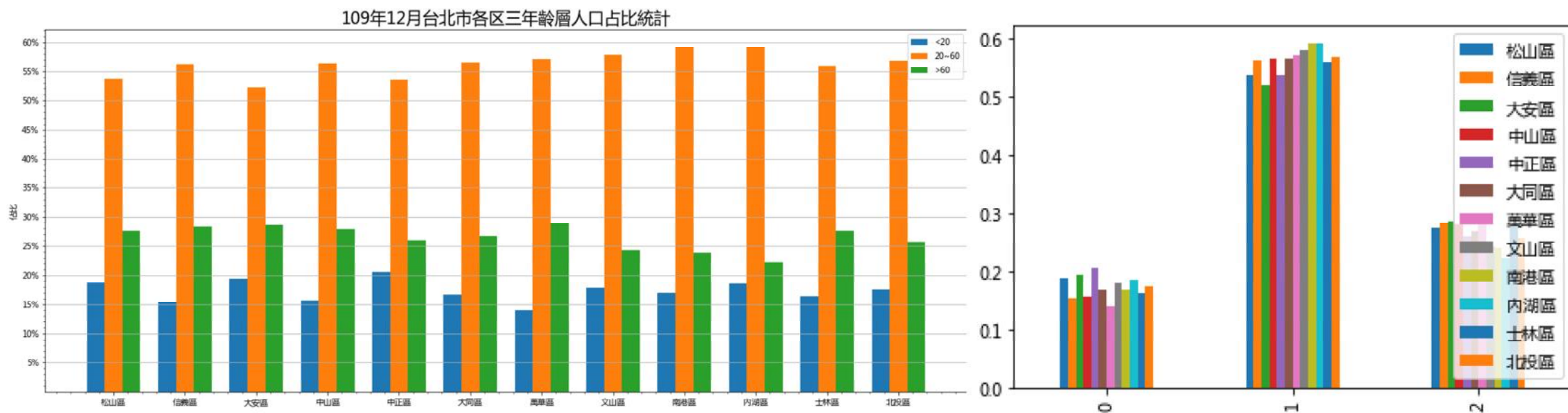


索引	松山區	信義區	大安區	中山區	中正區	大同區	萬華區	文山區	南港區	內湖區	士林区	北投區
101年	1206	1529	991	3274	855	2182	2039	3780	2144	4674	2392	2431
102年	185	628	-223	2547	1426	1441	1153	2548	1924	3593	1241	1828
103年	-405	812	33	1526	586	949	708	1902	1185	2658	732	1432
104年	-12	842	566	1639	-297	544	817	2444	1402	2510	502	1401
105年	-1074	-1143	-1465	200	-1367	-351	-459	770	689	1175	-66	-14
106年	-1617	-2083	-1772	-122	-1436	-973	-1377	-984	-115	-125	-1473	-1003
107年	-1394	-2082	-1581	-1212	-892	-1089	-1899	-560	-356	-309	-1604	-1154
108年	-1344	-2691	-821	-1571	-1169	-1492	-2318	-1089	-961	-603	-2528	-1474
109年	-2893	-4037	-3333	-2573	-1837	-2023	-3328	-3113	-1181	-2702	-4075	-2414

- 台北市自105年開始大部分行政區人數開始連續下滑

```
result9=dict()
temp=[]
for item9 in arealist:
    mask9=(df['性別']=="總計")&(df['月份']=="6月")&(df["行政區"]==item9)
    df9=df[mask9]
    result9[item9]=list(df9['總計'].diff().values)
    temp=df9['總計'].diff()
#tempdf=pd.DataFrame(list(zip(result[0], result[1])))
indexlist=list(df9['年份'].values)
dfnew=pd.DataFrame(result9,index=indexlist)
dfnew=dfnew.drop('100年',axis=0)
print(dfnew,'\n')
```


Matplotlib or Pandas Plot?





■ 謝謝聆聽

